

Universal Multiple-Octet Coded Character Set
International Organization for Standardization
Organisation Internationale de Normalisation
Международная организация по стандартизации

Doc Type: Working Group Document

Title: Proposal for encoding the Varang Kshiti script in the BMP of the UCS

Source: Michael Everson, EGT (IE)

Status: Expert Contribution

Action: For consideration by JTC1/SC2/WG2 and UTC

Date: 1999-01-29

A. Administrative

1. Title

Proposal for encoding the Varang Kshiti script in the BMP of the UCS.

2. Requester's name

Michael Everson, EGT (WG2 member for Ireland).

3. Requester type

Expert contribution.

4. Submission date

1999-01-29.

5. Requester's reference

6a. Completion

This is a complete proposal.

6b. More information to be provided?

No.

B. Technical -- General

1a. New script? Name?

Yes. Varang Kshiti.

1b. Addition of characters to existing block? Name?

No.

2. Number of characters

50

3. Proposed category

Category A.

4. Proposed level of implementation and rationale

Level 1 noncombining.

5a. Character names included in proposal?

Yes.

5b. Character names in accordance with guidelines?

Yes.

5c. Character shapes reviewable?

Yes (see below).

6a. Who will provide computerized font?

Michael Everson.

6b. Font currently available?

Yes.

6c. Font format?

TrueType.

7a. Are references (to other character sets, dictionaries, descriptive texts, etc.) provided?

Yes, see bibliography below.

7b. Are published examples (such as samples from newspapers, magazines, or other sources) of use of proposed characters attached?

No.

8. Does the proposal address other aspects of character data processing?

Yes, see Unicode properties below.

C. Technical -- Justification**1. Contact with the user community?**

No. We need to contact Norman Zide and get his opinion of this proposal.

2. Information on the user community?

Speakers of the Ho language, whose population is 1,026,00, with 25%–50% literacy, according to the SIL *Ethnologue*. The *Ethnologue* only mentions the Oriya and Devanagari script areas.

3a. The context of use for the proposed characters?

To write the Ho language. Latin, Telugu, and Oriya scripts have also been used to write Ho.

3b. Reference

See bibliography.

4a. Proposed characters in current use?

Yes.

4b. Where?

In primary and adult education (general use)

5a. Characters should be encoded entirely in BMP?

Yes.

5b. Rationale

Contemporary use.

6. Should characters be kept in a continuous range?

Yes.

7a. Can the characters be considered a presentation form of an existing character or character sequence?

No.

7b. Where?**7c. Reference****8a. Can any of the characters be considered to be similar (in appearance or function) to an existing character?**

No.

8b. Where?**8c. Reference****9a. Combining characters or use of composite sequences included?**

No.

9b. List of composite sequences and their corresponding glyph images provided?**10. Characters with any special properties such as control function, etc. included?**

No.

E. Proposal

The Varang Kshiti script is used to write the Ho language. Ho is a North Munda language, which family, together with the Mon-Khmer languages, makes up Austro-Asiatic. Varang Kshiti was devised by charismatic community leader Lako Bodra as part of a comprehensive cultural program, and was offered as an improvement over scripts used by Christian missionary linguists. Varang Kshiti is used in primary and adult education, and is published in a variety of printed materials.

Structure

Varang Kshiti is written from left to right, with the vowel characters in the order they are spoken; complex vowel placement and variant forms found in Brahmi-derived scripts does not obtain in Varang Kshiti.

Names and ordering

Varang Kshiti has ten vowel characters: simple vowels *a*, *i*, *u*, mixed vowels *e*, *o* (an analysis based on Sanskrit practice), and ligatures. Characters are arranged in rows of three, in a systematic arrangement not unlike the Devanagari arrangement. The letter H used after a vowel indicates vowel lengthening. Zide notes that there is a special symbol for the syllable *om*, but he does not show it. I have presented an *o* with *anusvara* as a placeholder for the time being. The last four of the vowels are described as “ligatures”; though *anusvara* + *anusvara* yields ’ (*apostrophe*, probably glottal stop), and *a* + *a* yields *y*. Long *i* and *u* are ligatures (Ɛ = ƒ+ƒ, ƚ = L+L), but long *a* is written Ɔ ʉ *ah*.

Digits and numbers

No zero is given in Zide 1996; the tens are used only in stand-alone position; 11 is coded 𑄛 1+1, 𑄛𑄛 is coded XX 2+3, 57 is coded 𑄛𑄛 5+7. The tens could be considered ligatures of the digits 1-9 plus a notional zero (so 𑄛 = 𑄛+𑄛, 𑄛 = 𑄛+𑄛). If this were the case (which would be an advantage in terms of computer processing (what happens in Tamil?)), the tens should not be encoded explicitly. This notional zero is unattested. Information on numbers greater than 99 are cannot be derived from Zide’s article.

Unicode Character Properties

Spacing letters, category “Lo”, bidi category “L” (strong left to right)

xx00 - xx1E, xx33

Numbers, decimal digits, category “Nd”, bidi category “L” (strong left to right)

xx20 - xx29

Numbers, other, category “No”, bidi category “L” (strong left to right)

xx2A - xx32

Bibliography

Zide, Norman. 1996. “Scripts for Munda languages”, in Peter T. Daniels and William Bright, eds. *The world’s writing systems*. New York; Oxford: Oxford University Press. ISBN 0-19-507993-0

Zide gives other sources in his bibliography, none of which I have seen.

TABLE XXX - Row xx: VARANG KSHITI

	xx0	xx1	xx2	xx3
0	᳚	᳛	᳜	᳝
1	᳞	᳟	᳠	᳡
2	᳢	᳣	᳤	᳥
3	᳦	᳧	᳨	ᳩ
4	ᳪ	ᳫ	ᳬ	
5	᳭	ᳮ	ᳯ	
6	ᳱ	ᳲ	ᳳ	
7	᳴	ᳵ	ᳶ	
8	᳷	᳸	᳹	
9	ᳺ	᳻	᳼	
A	᳽	᳾	᳿	
B	᳼	᳽	᳾	
C	᳿	᳾	᳼	
D	᳼	᳽	᳾	
E	᳽	᳾	᳿	
F	᳾		᳿	

G = 00
P = 00

